# SEGMENTASI PASAR KOPI MENGGUNAKAN METODE KNN DI INDONESIA

**Muhamad Maksum Hidayat[1)], Wahyu Nugroho[2)], dan Arief Setyanto[3)]**

[1,2,3)]MTI Universitas Amikom Yogyakarta

e-mail: maksum.hidayat24@gmail.com[1)], wahyunugraha1395@gmail.com[2)], arief_s@amikom.ac.id[3)]

**ABSTRAK**

*Kopi merupakan salah satu komoditas perkebunan yang banyak dikonsumsi di Indonesia dan memiliki peluang besar untuk dikembangkan, khususnya bidang pengembangan pasar. Hal yang menjadi kendala bagi pengusaha kopi adalah segmentasi pasar kopi yang tepat dengan kondisi negara Indonesia yang sangat luas. Penelitian ini bertujuan untuk mensegmentasi pasar kopi di Indonesia dengan memanfaatkan data gambar kopi yang ada di media sosial instagram dengan klasifikasi citra menggunakan algoritma K-Nearest Neighbor dan ekstraksi fitur histogram, sehingga dapat menentukan segmentasi pasar kopi yang tepat di Indonesia.*

***Kata Kunci**: Kopi, K-Nearest, Instagram, Histogram.*

**ABSTRACT**

*Coffee is one of the most widely consumed plantation commodities in Indonesia and has great opportunities to be developed, especially in the field of market development. The thing that becomes an obstacle for coffee entrepreneurs is the right segmentation of the coffee market with the vast Indonesian state. This study aims to segment the coffee market in Indonesia by utilizing coffee image data available on Instagram social media with image classification using the K-Nearest Neighbor algorithm and histogram feature extraction, so as to determine the exact coffee market segmentation in Indonesia.*

***Keywords**: Coffe, K-Nearest, Instagram, Histogram.*

## I. PENDAHULUAN

Kopi merupakan salah satu komoditas perkebunan yang banyak dikonsumsi di Indonesia dan mempunyai peluang besar untuk dikembangkan, khususnya bidang pengembangan pasar atau perdagangan. Peluang ini tentunya harus dapat ditemukan intensifikasi dan dimanfaatkan oleh produsen kopi baik oleh petani maupun produsen kopi yang bukan petani.

Tingkat konsumsi kopi di Indonesia selalu mengalami peningkatan setiap tahunnya seperti dikatakan dalam laporan *International Coffee Organitation* [1], bahwa Indonesia mengalami peningkatan konsumsi domestik sebesar 2,1% tiap tahunnya. Hal ini menjadi peluang besar bagi pengusaha kopi di Indonesia, namun yang menjadi kendala bagi pengusaha kopi adalah segmentasi pasar kopi yang tepat dengan kondisi negara Indonesia yang sangat luas.

Dizaman digital saat ini media sosial bisa dimanfaatkan sebagai media segmentasi pasar kopi, dikarenakan di media sosial banyak masyarakat pengguna media sosial mengupload aktivitas konsumsi kopi dari berbagai daerah di Indonesia. Dengan memanfaatkan data yang ada dari media sosial kita bisa memetakan konsumsi kopi di Indonesia, sehingga bisa membantu dalam menentukan pemasaran kopi di Indonesia.

Proses pengumpulan data dilakukan dengan mengambil gambar kopi dari media sosial instagram serta menggunakan hastag kopi dan daerah yang diinginkan. Dari gambar yang diperoleh maka perlu diklasifikasikan menjadi gambar kopi dan non kopi. Han [2] mengatakan bahwa klasifikasi merupakan teknik data mining yang digunakan untuk memprediksi kategori dari objek yang belum memiliki kategori. Salah satu metode yang dapat digunakan untuk mengklasifikasikan gambar adalah metode KNN.

KNN atau K-Nearest Neighbor adalah metode klasifikasi yang menentukan kategori berdasarkan mayoritas kategori pada *k*- tetangga terdekat. Jika D merupakan sekumpulan data pelatihan maka ketika data uji *d* disajikan, algoritma akan menghitung jarak antara setiap data dalam D dengan data uji *d*. Perhitungan jarak dilakukan dengan menggunakan *euclidian distance.*

Kemudian, *k* buah data dalam D yang memiliki jarak terdekat dengan *d* diambil. Himpunan *k* merupakan *k-nearest neighbor.* Selanjutnya, kategori *d* ditentukan berdasarkan mayoritas kategori dalam himpunan *k*-tetangga terdekat. Meskipun sederhana, KNN sering memberi hasil akurasi yang tinggi pada banyak kasus[3]. Selain itu metode KNN juga digunakan karena sangat fleksibel dan dapat bekerja dengan berbagai keputusan[4].

Setelah data diperoleh dan diklasifikasi yang menghasilkan nilai akurasi gambar yang diperoleh, maka akan dilakukan *rangking* terhadap nilai akurasi dataset yang ada. Penelitian ini bertujuan untuk mensegmentasi pasar kopi di Indonesia menggunakan data dari instagram dan metode klasifikasi KNN untuk membantu pengusaha kopi dalam menentukan segmen pasar kopi lebih akurat.

## II. STUDI PUSTAKA

### A. *Penelitian Terkait*

Berbagai penelitian mengenai klasifikasi gambar telah dilakukan sebelumnya. Perbedaan yang paling kentara dari setiap penelitian terletak pada dataset dan fitur-fitur yang digunakan. Perbedaan lain juga terletak pada metode klasifikasi yang digunakan.

Penelitian yang dilakukan Vailaya[5] menjelaskan bahwa pengelompokan gambar pada kategori menggunakan fitur visual tingkat rendah cukup menantang untuk kasus *content based image retrieval* (CBIR). Metode *bayesian* diterapkan dalam penelitian tersebut untuk mengetahui konsep dari fitur-fitur gambar tingkat rendah. *Output* dari setiap gambar uji akan masuk dalam kategori tertentu. Gambar yang diuji adalah gambar berkategori liburan yaitu libura *indoor* atau *outdoor.* Kelas *outdoor* dikategorikan lebih lanjut sebagai kota dan pemandangan. Kategori pemandangan kemudian dikategorikan lebih lanjut menjadi kategori matahari terbenam, hutan, dan gunung. Hasil penelitian ini menunjukkan bahwa akurasi dari kategori *indoor* dan *outdoor* adalah 90,5%, akurasi dari kategori kota dan pemandangan adalah 95,3%, dan akurasi dari kategori matahari terbenam, hutan, dan gunung adalah 96,6%.

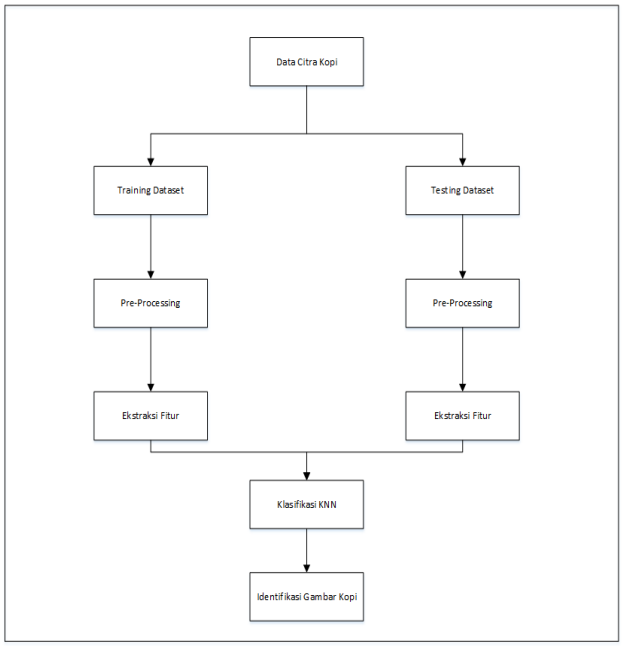
Eka[6] menjelaskan bahwa kebutuhan yang tinggi perusahaan manufaktur untuk menjaga kualitas produk membutuhkan kontrol pada akhir proses produksi. Kontrol inspeksi visual produk yang dilakukan manusia tidak sepenuhnya dapat diandalkan. Hal ini dapat diatasi dengan menggunakan visi komputer sebagai kontrol inspeksi visual produk. Komputer tidak dapat membedakan tekstur seperti halnya penglihatan manusia. Oleh karena itu, perlu digunakan analisis tektur untuk mengetahui pola suatu citra digital berdasarkan ciri yang diperoleh secara matematis. Penelitian ini menggunakan salah satu metode analisis tekstur yaitu metode *haar transform*. Ciri tekstur dapat diperoleh dari parameter panjang dan lebar garis dalam pixel. Kedua ciri tersebut kemudian digunakan untuk menentukan kategori menggunakan metode KNN. Keluaran dari sistem pengenalan dikelompokkan menjadi 3 kategori yaitu: pakan robek, pakan tercampur, dan pakan rangkap. Dari hasil yang diperoleh disimpulkan bahwa metode *haar transform* dapat digunakan untuk membedakan tekstur. Metode KNN diuji menggunakan k = 1,5,10, dan 20.

Herdiyani[7] dalam penelitiannya menjelaskan terkait penggunaan SVM untuk kasus *content based image retrieval* (CBIR). Pada penelitian ini, komponen warna digunakan sebagai fitur utama dan algoritma SVM dioptimalkan menggunakan *sequential minimal optimization* ( SMO. Eksperimen menggunakan *caltech dataset* dan beberapa gambar yang bersumber dari www.flower.vg. Hasil menunjukkan bahwa akurasi metode klasifikasi yang menggunakan jarak *euclidean* untuk mencari tetangga terdekat, yaitu 50,91%.

Perbedaan antara penelitian yang dilakukan dengan penelitian-penelitian sebelumnya terletak pada dataset dan fitur-fitur yang digunakan. Dataset yang digunakan pada penelitian ini adalah data gambar kopi yang diambil dari google image, dan untuk data testing diambil dari instagram dengan menggunakan hastag.

## III. METODE PENELITIAN

Metode penelitian identifikasi kopi menggunakan metode KNN dan ekstraksi fitur histogram adalah sebagai berikut:




Gambar 1. Metode yang diusulkan

### A. *Pengumpulan Data*

Pengumpulan data pada penelitian ini diambil dari media sosial. Media sosial didefinisikan sebagai istilah umum untuk interaksi sosial yang dibangun atas banyak media digital dan teknologi, yang memungkinkan pengguna untuk membuat dan berbagi konten dan bertindak secara kolaboratif[8][9]. Salah satu sosial media penyedia layanan interaksi sosial berbasis gambar dan text adalah instagram. Data yang digunakan dalam penelitian ini berupa citra gambar yang diambil dari instagram dengan kategori hastag (#) kopi dan hastag (#) daerah yang akan dijadikan objek penelitian, daerah tersebut adalah Yogyakarta, Jakarta, Medan, Surabaya, dan Bandung.

### B. *K – Nearest Neighbor*

Algoritma *K-Nearest Neighbor* (KNN) adalah metode yang menggunakan algoritma *supervised* dimana hasil *query instance* diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Algoritma KNN bertujuan untuk mengklasifikasikan objek-objek yang baru menurut atribut dan *training sample*[10].

### C. *Ekstraksi Fitur Histogram*

Histogram adalah metode yang digunakan untuk mendapatkan tekstur dengan dasar pada *histogram. Histogram* citra merupakan grafik yang menggambarkan penyebaran pada nilai-nilai intensitas piksel suatu citra atau bagian yang tertentu pada suatu citra. Dari segi *histogram* dapat diketahui bahwa frekuensi kemunculan nisbi (*relative*) dari suatu intensitas pada citra tersebut. Metode *histogram* merupakan metode statis order satu untuk mendapatkan fitur-fitur tekstur[11].

Fitur-fitur yang dapat diketahui dari metode *histogram* antara lain yaitu: rerata intensitas, deviasi standar, *skewness,* energi, entropi, dan kehalusan (*smoothness*). Persamaan pada perhitungan ke-6 ciri tersebut berbasis *histogram* dijelaskan dibawah ini:

Fitur pertama adalah rerata intensitas dengan rumus sebagai berikut:

$$m = \sum_{i=0}^{L-1} i.p\,(i) \quad (1)$$

Dalam hal ini, *i* merupakan aras keabuan pada citra *f* dan *p(i)* merupakan probabilitas kemunculan *i* dan L merupakan nilai aras keabuan tertinggi. Rumus yang diatas akan menghasilkan rerata kecerahan pada citra.

Fitur yang kedua yaitu deviasi standar, adapun rumusnya adalah sebagai berikut:

$$\sigma = \sqrt{\sum_{i=0}^{L-1}(i-m)^2\,p(i)} \quad (2)$$

Pada rumus deviasi standar, $\sigma^2$ dinamakan varians atau momen orde dua tenomalisasi dikarenakan *p (i)* adalah fungsi peluang. Fitur deviasi akan memberikan ukuran kekontrasan.

Pada fitur *skewness* merupakan ukuran ketidaksimetrisan terhadap rerata intensitas. Adapun rumus *skewness* adalah sebagai berikut:

$$Skewness = \sum_{i=1}^{L-1}(\,i-m\,)^3\,p(i) \quad (3)$$

*Skewness* juga sering disebut dengan momen order tiga ternomalisasi dimana nilai negatif menyatakan bahwa distribusi pada kecerahan lebih condong ke kanan terhadap rerata. Pada praktiknya, nilai *skewness* dibagi menjadi $(L\text{-}1)^2$ agar lebih ternomalisasi.

Fitur ke empat adalah fitur energi yang merupakan ukuran yang menyatakan distribusi inetensitas piksel pada jangkauan aras keabuan. Rumus fitur energi adalah sebagai berikut:

$$energi = \sum_{i=0}^{L-1}[\,p(i)]^2 \quad (4)$$

Citra yang sama dengan satu nilai aras keabuan akan mempunyai nilai pada energi yang maksimum, yaitu sebesar 1. Secara umum, citra dengan sedikit aras keabuan akan memiliki energi yang lebih tinggi dibanding yang memiliki banyak nilai aras keabuan. Energi sering disebut juga sebagai keseragaman.

Fitur selanjutnya adalah fitur entropi. Fitur entropi mengindikasikan kompleksitas citra. Rumus fitur entropi adalah sebagai berikut:

$$Entropi = \sum_{i=0}^{L-1} p(i) Log_2 (p(i)) \qquad (5)$$

Semakin tinggi nilai entropinya maka akan semakin kompleks citra tersebut. Seperti diketahui, entropi serta energi berkecenderungan berkebalikan. Entropi juga merepre-sentasikan jumlah informasi yang terkandung pada sebaran data.

Pada rumusan ekstraksi fitur histogram yang terakhir yaitu fitur kehalusan (*smothness*) merupakan fitur untuk mengukur tingkat kehalusan atau kekasaran pada intensitas citra. Penjabarannya adalah sebagai berikut:

Pada rumusan diatas, $\sigma$ mrupakan deviasi standar. Berdasarkan rumus yang disebutkan diatas, nilai pada R yang rendah menunjukkan bahwa citra memiliki intensitas yang kasar.Seperti yang diketahui, di dalam menghitung *smoothness,* varians perlu dinormalisasi sehingga nilainya berada di dalam jangkauan [0-1] dengan cara membagi $(L\text{-}1)^2$.

### *D. Jarak Minkowski*

Jarak *Minkowski* adalah salah satu perhitungan jarak pada metode K-Nearest Neighbor. Prinsip perhitungan *minkowski* adalah mengukur jarak v1 dan v2 yang merupakan dua vektor yang jaraknya dihitung dan N yaitu menyatakan panjang dari vektor. Sebagai contoh, dengan dua vektor yang sama dengan di depan ($v_1$ =[4,3,6] ) dan $v_2$ =[2,3,7], jarak *Minkowski* kedua vektor tersebut untuk *p* berupa:

$$jv_1, jv_2 = \sqrt[1/p]{\sum_{k=1}^{n} |v_1(k) - v_2(k)|p}$$

## IV. HASIL DAN PEMBAHASAN

### *A. Data Citra*

Data citra atau gambar yang digunakan berupa gambar yang didapatkan dari hasil *crawling* gambar dimedia sosial instagram dengan filter hastag (#) kopi dan hastag (#) daerah di Indonesia. Daerah yang dipilih untuk penelitian ini adalah Yogyakarta. Jakarta, Medan, Surabaya, dan Bandung, dengan masing-masing 20 citra setiap hastag daerah. Sehingga citra yang diperoleh berjumlah 100 citra. Dari jumlah 100 citra 80 data citra kopi digunakan sebagai data *training* (data latih) serta 20 data citra digunakan sebagai data *testing* (data uji), data yang diujikan bergantian 20 citra setiap kategori daerahnya.

Langkah selanjutnya yaitu pengelompokkan dari 100 dataset yang dibagi menjadi 2 kategori yaitu 20 *data test* dan 80 *data training*. Dari 100 dataset tersebut , 20 data citra digunakan sebagai data test sedangkan 90 data citra digunakan sebagai data *training*.

Data citra kopi yang digunakan sebagai data *train* adalah sebagai berikut:

Tabel 1. Data *training*

| NO | Citra | Kelas |
|---|---|---|
| 1 | 1.jpg | Kopi |
| 2 | 2.jpg | Kopi |
| 3 | 46.jpg | Non kopi |
| 4 | 56.jpg | Non kopi |

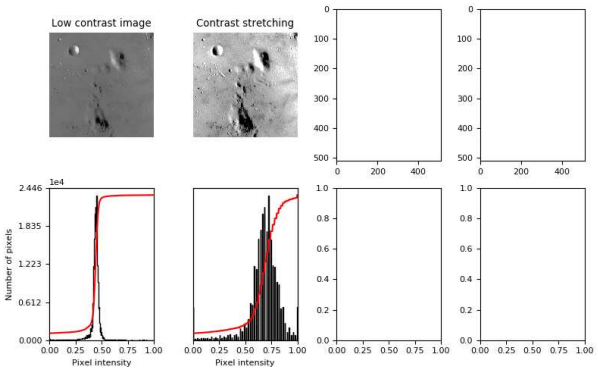
Sedangkan data *testing* (data uji) adalah sebagai berikut:

Tabel 2. Data *testing*

| NO | Citra | Kelas |
|---|---|---|
| 1 | 1.jpg | Kopi |
| 2 | 2.jpg | Kopi |
| 3 | 8.jpg | Non kopi |
| 4 | 12.jpg | Non kopi |

### *B. Ekstraksi Fitur Histogram*

Pada ekstraksi fitur menggunakan metode *histogram* terdapat 6 fitur yang diantaranya adalah: rerata intensitas, deviasi, *skewness,* energi, entropi, dan kehalusan (*smoothness*). Pada Gambar dibawah ini merupakan hasil ekstraksi fitur *histogram* dari 100 data citra yang terbagi menjadi dua kategori yaitu kopi dan non kopi. Citra tersebut kemudian diproses menggunakan ekstraksi fitur histogram, dari hasil pengolahan tersebut.



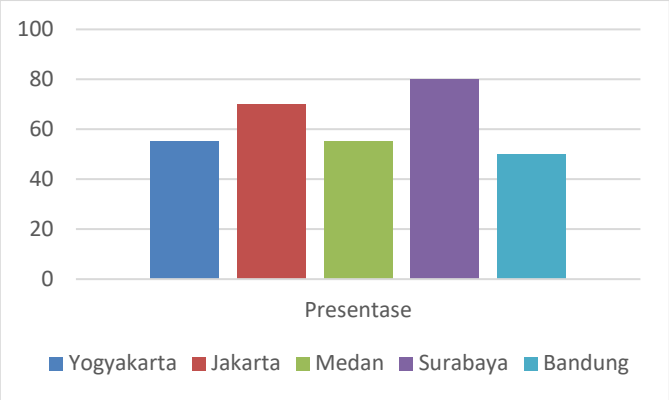

Gambar 2. Ekstraksi Fitur Histogram

Data diatas merupakan beberapa data citra kopi yang sudah diekstraksi fitur mengggunakan ekstraksi fitur

histogram yang menghasilkan angka-angka piksel yang nantinya hasil dari angka-angka piksel tersebut akan diklasifikasi kedalam algoritma K-Nearest Neighbor.

### C. *Algoritma K – Nearest Neighbor*

Dari hasil proses klasifikasi menggunakan algoritma KNN dengan data *train* dan data *test* serta menggunakan K=3 adalah sebagai berikut:




Gambar 3. Hasil Klasifikasi Citra Kopi

Berdasarkan gambar 3 diatas menunjukkan bahwa peminat kopi paling tinggi di Indonesia dari 5 kategori daerah yang dijadikan objek data testing Surabaya menjadi kota tertinggi peminat kopinya dengan prosentase 80%, kemudian Jakarta dengan prosesntase 70%, Medan 55%, serta Yogyakarta dan Bandung dengan prosesntase 50%. Akurasi yang diperoleh dalam klasifikasi citra kopi menggunakan algoritma K-Nearest Neighbor ini sebesar 83,3%.

## V. KESIMPULAN

Berdasarkan hasil dari penelitian tentang segmentasi pasar kopi menggunakan algoritma K-Nearest Neighbor dengan ekstraksi fitur histogram, dapat disimpulkan beberapa hal berikut:

1. Pada identifikasi citra kopi menggunakan algoritma *K-Nearest Neighbor* dengan menggunakan 100 data citra yang terdiri dari gambar kopi dan non kopi. Data uji sebanyak 20 citra dan memperoleh akurasi sebesar 83,3% .
2. Hasil dari segmentasi pasar kopi dengan data yang diambil dari instagram berdasar hastag (#) kopi dan hastag (#) daerah menghasilkan bahwa peminat kopi tertinggi adalah Surabaya.
3. Berdasarkan hasil diatas dapat disimpulkan bahwa dengan menggunakan algoritma *K-Nearest Neighbor* serta menggunakan ekstraksi fitur *histogram* dapat diterapkan dengan baik pada identifikasi citra kopi dan non kopi

## DAFTAR PUSTAKA

[1] International coffee Organization, “Domestic consumption by all exporting countries In thousand 60kg bags,” *Int. Coffee Organ.*, pp. 2–4, 2017.

[2] J. Han and M. Kamber, “Data Mining : Concepts and Techniques ( 2nd edition ) Bibliographic Notes for Chapter 5 Mining Frequent Patterns , Associations , and Correlations,” *Quest*, 2006.

[3] K. M. Laina Farsiah, Taufik Fuadi Abidin, “Klasifikasi Gambar Berwarna Menggunakan K-Nearest Neighbor dan Support Vector Machine,” *SNASTIKOM*, 2013.

[4] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, dan Usage Data*. Berlin: Springer Berlin Heidelberg.

[5] A. Vailaya, A. Member, M. A. T. Figueiredo, and A. K. Jain, “Image Classification for Content-Based Indexing,” vol. 10, no. 1, pp. 117–130, 2001.

[6] E. Nurmajaya, “Aplikasi Pengolahan Citra DIgital Dalam Klasifikasi Cacat Kain Grey Menggunakan Metode K-Nearest Neighbor,” UII Yogyakarta, 2011.

[7] Herdiyani, Y, and Ah. Buono, “Klasifikasi Citra Dengan Support Vector Machine Pada Sistem Temu Kembali Citra,” in *Seminar Nasional Sistem dan Informatika*, 2007.

[8] D. Harjowinoto, A. Noertjahyana, and J. Andjarwirawan, “Vulnerability Testing pada Sistem Administrasi Rumah Sakit X,” *J. Infra*, vol. 4, no. 1, p. pp.227-p.232, 2016.

[9] D. Schoder, P. A. Gloor, and P. T. Metaxas, “Social Media and Collective Intelligence—Ongoing and Future Research Streams,” *KI - Künstliche Intelligenz*, vol. 27, no. 1, pp. 9–15, 2013.

[10] M. I. Sikki, “Pengenalan Wajah Menggunakan K-Nearest Neigbor Dengan PraProses Transformasi Wavelet,” *J. Paradig.*, vol. X No.2, 2009.

[11] S. Q. Nugroho and R. A. Pramunendar, “Pengelompokan Kayu Kelapa Menggunakan Algoritma K-Means Berdasarkan Tekstur Citra Kayu Kelapa Dua Dimensi ( 2D ),” *Univ. Dian Nuswantoro*, pp. 1–5, 2015.